On the Dimensionality of Wireless Connectivity Traces

George Roussos
Birkbeck College
University of London
g.roussos@birkbeck.ac.uk

Abstract

Routing efficiency in wireless networks can be greatly improved by matching mobile host connectivity patterns. To this end, over the past few years considerable effort has been invested in developing predictors of mobility patterns that is, models of mobile host movement so that for a specific sequence of recorded locations to predict the most likely subsequent location. In this paper, I initiate a study of the structure of the connectivity space itself. I analyze samples from the Dartmouth data set[1] *and conduct an exploratory study of the structure of the underlying space. In particular, I compute the singular values of the time-weighted connectivity matrix, and relate this result to principal component analysis. Initial findings indicate that the degrees of freedom of the space induced by the connectivity matrix is very low with respect to the number of access points and mobile hosts involved; that the dimensionality of the underlying space grows slowly with the size of the sample; and, that the distribution of its eigenfrequencies follows a power law.*

1. Introduction

Wireless networks can better serve mobile hosts by employing client location information to better anticipate connectivity patterns. Predicting accurately the location of hosts can potentially improve the performance of wireless routing and the robustness of the network infrastructure itself, thus improving the user experience for a variety of applications. These improvements lead to a better user experience, to a more cost-effective infrastructure, or both. As a result, during the past few years a number of location predictors have been proposed in the literature based on a variety of complementary techniques including Markov-based, compression-based, PPM, and SPM mechanisms. Such approaches infer models of mobile host movement patterns so that for a specific recorded sequence of recorded locations to predict the most likely subsequent location. A comprehensive comparative evaluation study of the relative performance of several of these algorithms on the Dartmouth mobile data trace [2] including a detailed description of their structure and performance can be found in [4].

In this poster, rather than focusing on predictors I initiate the study of the structure of the connectivity space itself so as to understand its core characteristics. To do this, I also employ the Dartmouth data trace to reconstruct the time-weighted connectivity matrix between access points and mobile clients which we use as the basis of this investigation. Several techniques are employed to explore the structure of this space, with a view in all cases to identify the existence of a small subset of components or eigenfrequencies that characterize accurately the client connectivity behavior. Effectively, I aim to determine a basis for a low-dimensional projection of the connectivity matrix that provides an appropriately accurate approximation to the overall connectivity patterns.

2. SVD and Principal Component Analysis

Let $L = \{l_{ij}\}$ be the $m \times n$ connectivity matrix of a data trace defined by setting the element l_{ij} to be

[1] Many thanks too David Kotz and other members of the Dartmouth Centre for Mobile Computing for providing access to this data.

proportional to the time mobile host j is connected to access point i, for a data sample that includes traces of n hosts and m base stations. Note that row i of L describes the time each sighted client has spend connected to base station i, and column j of L describes the time spend by host j connected to each base station. Hence, I refer to the row vectors of L as the connectivity profile for access point i and to the column vectors as the connectivity pattern of client j. Due to the rather limited mobility of hosts in the Dartmouth traces, the matrix L is sparse.

The singular value decomposition (SVD) of L is

$$L = USV^T, \tag{1}$$

where U is an $m \times n$ matrix, S is an $n \times n$ diagonal matrix, and V also an $n \times n$ matrix. The columns of U are called the left singular vectors and form an orthonormal basis for the connectivity patterns of clients, that is $u_i u_j$ is zero except where $i = j$ when it is one. The rows of V^T contain the elements of the right singular vectors and form an orthonormal basis for the connectivity profiles for the access points.

The matrix S is zero everywhere except at the diagonal, that is $S = \mathrm{diag}(s_1, \ldots, s_n)$, where it contains the so-called singular values s_k of L. Singular values are ordered so that the highest singular value is in the upper left index of the matrix S. Note that if L is square, that is $m = n$, then the SVD is equivalent to the solution of the eigenvalue problem.

With the SVD at hand, we can compute the closest r-rank matrix to L as follows

$$L^{(r)} = \sum_{k=1}^{r} u_k s_k v_k^T \tag{2}$$

so that $L^{(r)}$ minimizes the sum of the squares of the difference of the elements of L and $L^{(r)}$. Standard approaches to compute the SVD can be found in [3].

2.1 Relation to principal component analysis

Principal component analysis (PCA) captures the variance in a dataset in terms of its so-called principle components. The SVD is intimately related to PCA when principal components are calculated using the covariance matrix. If each column of L is centered then $L^T L$ is proportional to the covariance matrix of the connectivity profile for the access points. Moreover, diagonalisation of $L^T L$ yields V^T and thus the principal components of the connectivity profile. in other words, the right singular vectors of L are the same as the required principal components. Further, the eigenvalues of $L^T L$ are the singular values of L, which are proportional to the variances of the principal components. Overall, the matrix US contains the principal component scores, which are the coordinates of the connectivity profile in the space of principal components. If instead of centering the columns of L we center its rows then $L^T L$ is proportional to the covariance matrix of the connectivity patterns of the mobile clients. Similar to above, the left singular vectors are also the principal components of the connectivity pattern space; the singular values are proportional to the variances of the principal components; and the matrix SV^T contains the principal component scores, that is the coordinates of the connectivity patterns in the space of principal components.

2.2 Results

In this section I report on preliminary results of the analysis of connectivity patterns using the Dartmouth mobile data trace. Logfiles provided by the CMC were post-processed to extract the connection/disconnection patterns of particular hosts to particular base stations. Several data sets where developed consisting of up to $90,000$ samples, which were subsequently analyzed following the discussion in previous sections.

As noted earlier the main focus of the analysis is to better understand the underlying structure of this space which in this case is characterized by the range of the singular values. For the largest sample we find that the majority of the singular values are relatively small with respect to the largest components, in fact only 4 of those are within 10% of the magnitude of the largest one. Figure 1 provides some more information regarding the actual distribution of the spectrum: the magnitude of the computed singular values appears to follow a power law distribution.

Both observations point towards the fact that it should be possible to reconstruct the full connectivity matrix using only a small number of principal components within a high accuracy. Indeed, using only the 29 principal components it is possible to achieve accuracy of the order of 0.1% for this sample. Thus, the variability of the connectivity patterns in the data set is very low and can be predicted very well using a subspace of very low dimensionality.

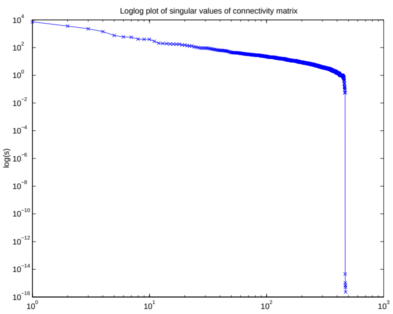

Figure 1. Logarithmic plot of the singular values of the connectivity matrix by size ($90,000$ samples).

3. Discussion and Conclusions

Several papers on wireless networking measurement observe that mobility patterns of individual hosts are likely to contain a considerable amount of periodicity. The underlying reasons for this are mainly economic and demographic, and dictate that clients move within a small range of different velocities, and travel along similar routes with most journeys starting and ending at similar places. As a consequence, there is only a small number of approximately discrete frequencies characterizing the behavior and the connectivity profile of the mobile station (and its user). In this paper I provide some preliminary results regarding the dimensionality of the space induced by the mobility patterns observed within a particular experimental setting. Moreover, some early results are highlighted regarding the number of degrees of freedom that exist in the data and have estimated the order of accuracy of reconstruction using a low-dimensional approximation.

This evidence can be used to potentially improve the performance of wireless routing and the robustness of the wireless network infrastructures. For example, using the computed principal components it is possible to pinpoint bottlenecks in wireless networks as well as loss points for example due to interference. Knowledge of such locations allows to locate additional resources where they are needed to improve reliability for example hop-by-hop rather than end-to-end packet loss recovery.

More importantly, these findings provide evidence in support of a recent conjecture by Jon Crowcroft [1] regarding the feasibility of a common network architecture that overarches both mobile wireless mesh and fixed networks. He observes the following correspondence between wireless mesh and wireline fixed networks:

Mesh network		Wireline network
Mobility	$\Longleftrightarrow$	Buffering
Freq. distribution		Router out degree

A critical element for the proposed unified reference model is that the distribution of journey frequencies follows a similar pattern to the distribution of communication popularity, as recorded in the out degree of the connectivity graph of Internet routers for example. This would imply that the routing system for both types of networks would have similar properties.

In this paper I provide a strong indication that this is indeed the case, since the computed singular values of the connectivity matrix or else the journey frequencies clearly appear to follow a power law distribution (cf. Figure 1) and is thus qualitatively similar to the router out degree on the Internet.

References

[1] J. Crowcroft. Communication to the UCL Mobile Systems Group. 10 September 2004.

[2] D. Kotz and K. Essien. “Analysis of a Campus-wide Wireless Network”. *Wireless Networks*, 11:115-133, 2005.

[3] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.

[4] L. Song, D. Kotz, R. Jain, X. He. “Evaluating next-cell predictors with extensive Wi-Fi mobility data”. *Proc. 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*:1414-1424, 2004.